Bivariate Data – Grade 11

Ohio Standards Connection

Data Analysis and Probability

Benchmark A
Create and analyze tabular and graphical displays of data using appropriate tools, including spreadsheets and graphing calculators.

Indicators
4. Create a scatter plot of bivariate data, identify trends, and find a function to model the data.
5. Use technology to find the Least Squares Regression Line, the regression coefficient, and the correlation coefficient for bivariate data with a linear trend, and interpret each of these statistics in the context of the problem situation.
8. Analyze and interpret univariate and bivariate data to identify patterns, note trends, draw conclusions, and make predictions.

Benchmark B
Use descriptive statistics to analyze and summarize data, including measures of center, dispersion, correlation and variability.

Lesson Summary:
Working together, students determine a regression line and a coefficient of correlation for bivariate data. Students determine relationships between the regression line and the coefficient of correlation in a group learning activity. Students examine scatter plots with similar lines of regression and different correlation coefficients to explore the value of correlation coefficients.

Estimated Duration: *90 minutes*

Commentary:
This lesson provides a review discussion of correlation versus causation as part of the pre-assessment. The lesson also provides the opportunity for students to learn the material in different ways, including visually, and symbolically. The lesson gives options for flexible use of time and gives an excellent introduction on lines of regression.

Pre-Assessment:
- Pair students and distribute *Getting Started,* Attachment A. This activity is a review of making a scatter plots and estimating a line of best fit.
- Discuss the students' answers for the line of best fit. Ask students why different pairs may have slightly different equations for the line of best fit. Discuss the decisions each pair made when determining which points to use to write the line equation and how those choices affected the equation.
- Discuss the students' graphs in Part Two of Attachment A, *Getting Started.*
- Discuss with students how as one variable increases (i.e. time spent on a diet) and the other decreases (i.e. weight), a negative correlation occurs. If one variable increases as the second one also increases, then a positive correlation occurs.

Instructional Tip:
Use uncooked spaghetti for this activity. If this is not available, use clear rulers.

Scoring Guidelines:
Informally observe students' progress and provide intervention as necessary.

Post-Assessment:
Allow students a few minutes to measure each other's arm and index finger lengths to the nearest tenth of a centimeter. Write the information as ordered pairs, (arm length, finger length), in a

Ohio Standards Connection

Indicator 8
Analyze and interpret univariate and bivariate data to identify patterns, note trends, draw conclusions, and make predictions.

Mathematical Processes

Benchmark C.
Assess the adequacy and reliability of information available to solve a problem.

location visible to all students, and direct students to plot each ordered pair on a coordinate plane.

- Instruct students to use technology to determine the regression equation and calculate the coefficient of correlation for the data.
- Ask students to describe a relationship between a person's arm length and index finger length using the calculated data.

Instructional Tip:
Depending on the class size, additional data can be collected from another class to make the relationship more obvious.

- Give students the following scenario:
  The Jolly Green Giant has joined our class. Only his finger could be measured; it was two meters in length.
- *Use the information calculated earlier to predict his arm length.*
- Write a statement about the reliability of this prediction, using the coefficient of correlation.

Scoring Guidelines:
Evaluate students' work using the rubric below.

3	• Creates a scatterplot correctly • Calculates the line of best fit, equation of line and the coefficient of correlation correctly • Describes the relationship between the two variables • Writes a prediction and a statement about the reliability of this prediction for the scenario using his/her information
2	• Creates a scatterplot correctly • Calculates an equation for the line of best fit as well as the coefficient of correlation, but one or both contain minor errors • Does not clearly describe the relationship between the two variables • Makes an accurate prediction and a statement about the reliability of this prediction for the scenario based on the flawed regression equation and/or correlation coefficient
1	• Includes a scatterplot • Calculates the line of best fit and its equation as well as the coefficient of correlation incorrectly • Does not clearly describe the relationship between the two variables • Writes a prediction and a statement about the reliability of this prediction for the scenario that does not follow from the inaccurate regression equation or correlation coefficient

Instructional Procedures:

Part One

1. Administer the pre-assessment.
2. Ask students what they know about bivariate data. Some students may analyze the prefix of the word (bi) and the stem (variate) to conclude that bivariate data measure two things that change in relationship to each other. Some students may remember discussing relationships such as weight and age or speed and distance.
3. Create an overhead transparency for Attachment B, *Putting It Together*, and display both graphs for students. Have students make conjectures about measurements that could be represented by the data. As students offer suggestions of possible representations, discuss whether or not those variables would have a cause and effect relationship or a coincidental relationship.
4. Assign students to small groups of three or four students. Have them plot the points on Attachment C, *Group Activity.* Have each group provide a story for each set of data and identify the variables in those situations.
5. Ask groups to share with the class the stories they created and the variables that go with those stories. As a class, select one story to represent each graph. On a transparency of the graphs, write the variables and a brief description of each situation. Discuss the relationships evident in the graphs and decide as a class how strong those relationships are (e.g., not very well related; appears very predictable; as one goes up, the other seems to go down).

Instructional Tip:

Allow the use of informal language at this point in the mathematical discussion. The formal language and explanations will make more sense to the students later in the lesson, once they have already understood the ideas in everyday language.

6. Explain to students that formal methods of finding the line of best fit and a number to represent the strength of the relationship between variables exists. Provide some explanation about the range of correlations coefficients (closer to 1 being stronger, zero meaning there is no relationship). However, do NOT at this point discuss the range going from -1 to 1. This will be discovered by the students later in the lesson. Ask students to brainstorm situations whose correlation coefficient would be closer to zero or closer to 1.
7. Demonstrate how to use technology (computer software or graphing calculator) to find the line of best fit and the correlation coefficient for the data on Page One of Attachment B, *Putting it Together.* Then have students work in pairs to find the equation and correlation coefficient for page two of Attachment B. Ask questions to summarize findings such as:
   - Ask students to compare the two problems.
   - How are they alike?
   - How are they different?
   - Could they predict which problem would have a higher coefficient of correlation by looking at the graphs?
   - What characteristics do they think make the correlation stronger (closer to one)?

Instructional Tips:

- If students are already familiar with using technology to find the line of best fit and coefficient of correlation using technology, divide them into groups to do so. If students are unfamiliar with the technology, use the demonstration to help students become familiar with the features of the calculator or computer software being used. Providing a list of steps (buttons to push or features to select) is helpful.
- Refer to the user's manual for instructions on graphing scatter plots and calculating the regression equation and correlation coefficient with a graphing calculator or computer.

8. Have students share with a partner something new they learned about correlation coefficients and lines of best fit today. Also have them share a question they still have about either topic.

Part Two

9. Ask students to state in one sentence what Part One of the lesson covered. Have students share the questions they still have about correlation coefficients or lines of best fit.
10. Divide the students into small groups of three or four. Ask students to refer to Attachment C, *Group Activity*, used in Part One of the lesson. Have the students work in their groups to make predictions about what the correlation coefficient might be for each graph. Share and record the groups' predictions.
11. Divide students into small groups of three to four students. Assign each group two problems from Attachment C, *Group Activity,* used in Part One of the lesson. Each group uses technology to create a scatter plot, find the line of best fit and calculate the coefficient of correlation for their assigned problems. Each group draws a scatter plot and records the coefficient and line equation for each problem on chart paper or a poster to display to the other groups.
12. Assign problems to at least two different groups, so each group can confirm the work on each problem with a different group. Provide time for groups to compare their results.

Instructional Tip:

Stress the importance of recording the data in a manner that others will be able to read clearly, when it is displayed

13. Instruct the groups to display their information after they have compared matching problems with another group.
14. Instruct students to circulate around the room, looking at the results for each problem. Allow about 10 minutes for viewing of results, and then have students return to their groups.
15. Ask students to discuss the following questions in their groups:
    - What do you notice about the regression equations?
    - How were the correlation coefficients similar or different?
    - What relationship or connection do you see between the slope of the line and the correlation coefficient?
    - How could a person tell which graphs represent a high correlation by looking at the scatter plots?
    - What does it mean to have a high or low correlation?

16. Discuss these questions as a class. Be sure during the discussion that students talk about why a negative correlation occurs, how as one variable increases (i.e. time spent on a diet), the other decreases (i.e. weight). Also emphasize that correlation does not imply causation. For example, the age a child begins to crawl may correlate with how well the child reads, but one does not actually cause the other.
17. Instruct students to create an example of two types of data that have similar numbers but different correlation coefficients. Each group should provide a scenario to describe and share the information with the class.

Differentiated Instructional Support:
Instruction is differentiated according to learner needs, to help all learners either meet the intent of the specified indicator(s) or, if the indicator is already met, to advance beyond the specified indicator(s).

- Students who need more practice with the technology used in this lesson may need detailed instructions to follow or more examples to help them remember the steps needed to create scatter plots, lines of best fit, and correlation coefficients.
- Students demonstrating evidence of understanding of the concepts may locate data in the Internet, or do a mini-research project to collect data to analyze, present as a scatter plot and determine the correlation.
- Students demonstrating evidence of understanding of the concepts may be interested in finding the line of best fit by hand. Many statistics textbooks and other resources provide instructions for doing the calculations.

Extensions:

- Show students the examples from *Putting It Together,* Attachment B. Graph the line of best fit found using technology. Demonstrate finding the distance from a data point to the line (draw vertical lines connecting each data point to the line of best fit). Find the length of each vertical line. Ask students how this might relate to the coefficient of correlation. Discuss how the numbers are related.
- Discuss with students research applications of correlation coefficients and how researchers use correlations to make decisions. For example, when testing a new medical treatment, is a correlation of .786 acceptable? What are typical correlation coefficients in medical research?
- Research as a class how to create a Least Squares Regression Line.

Home Connections:
Encourage students to read a newspaper or magazine and find a table or scatter plot involving bivariate data.

Materials and Resources:
The inclusion of a specific resource in any lesson formulated by the Ohio Department of Education should not be interpreted as an endorsement of that particular resource, or any of its contents, by the Ohio Department of Education. The Ohio Department of Education does not endorse any particular resource. The Web addresses listed are for a given site's main page, therefore, it may be necessary to search within that site to find the specific information required for a given lesson. Please note that information published on the Internet changes over time,

therefore the links provided may no longer contain the specific information related to a given lesson. Teachers are advised to preview all sites before using them with students.

For the teacher: grid paper, uncooked spaghetti or clear rulers, scissors, envelopes, graphing calculator or computer with software for scatter plots, regression equations and coefficients of correlation, and a transparency of Attachment B

For the student: copies of Attachment A, uncooked spaghetti or clear rulers, a graphing calculator or computer with software capable of performing required calculations, *Group Activity*, Attachment C (pre-cut for the activity), paper or flip chart paper for displaying data, and grid paper

Vocabulary:
- bivariate data
- coefficient of correlation/correlation coefficient
- collinear
- line of best fit
- line of regression
- linear
- scatter plot

Technology Connections:
- Technology is used in this lesson to calculate the regression lines and coefficient of correlation, and to create scatter plots.
- Students use the Internet to find information about bivariate data and correlation coefficients.

Research Connections:
Arter, Judith and Jay McTighe. *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Thousand Oaks, Calif: Corwin Press, 2001.

General Tips:
Prepare the group activity before class. Cut the problems and directions for student use.

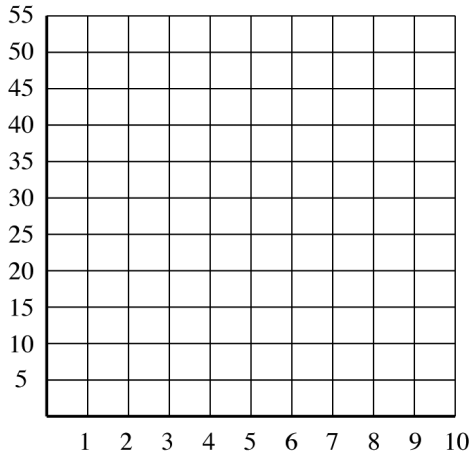
Attachments:
Attachment A, *Getting Started*
Attachment B, *Putting It Together*
Attachment C, *Group Activity*
Attachment D, *Group Activity Answer Key*

Attachment A
Getting Started

Name_____ **Date**_____

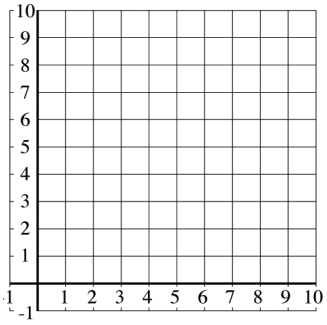
Part One
Directions: The table below details the number of students in an after school program over a period of ten weeks. Use the information to draw a scatter plot.

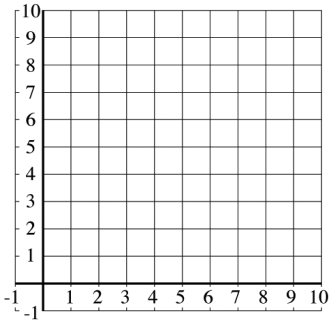
Time	Number
1	10
2	15
3	20
4	25
5	30
6	33
7	38
8	42
9	49
10	55

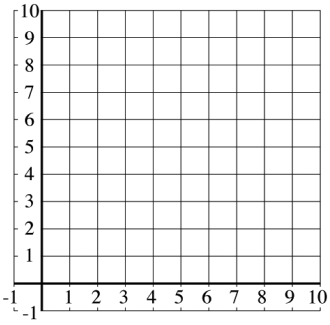


The points are not collinear. Place a piece of spaghetti so it fits the data as closely as possible. Use two points from the line (your spaghetti) to write the equation for that line.

Part Two
Directions: Sketch a graph to represent how each of the following may look on a graph
a. time spent on a diet and weight
b. age of a car and its value
c. amount of education and salary







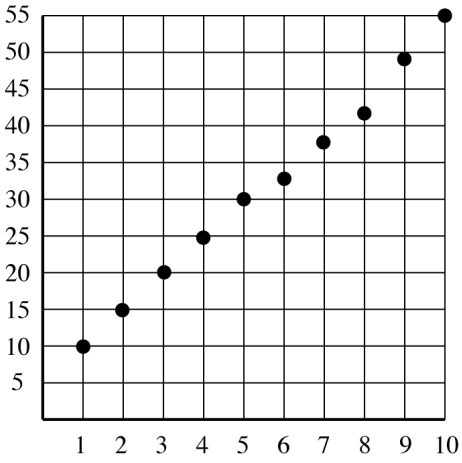
Attachment A
Getting Started Answers Key

Name_____ **Date**_____

Part One

Directions: The table below details the number of students in an after school program over a period of ten weeks. Use the information to draw a scatter plot.

Time	Number
1	10
2	15
3	20
4	25
5	30
6	33
7	38
8	42
9	49
10	55



The points are not collinear. Place a piece of spaghetti so it fits the data as closely as possible. Use two points from the line (your spaghetti) to write the equation for that line.

Part Two

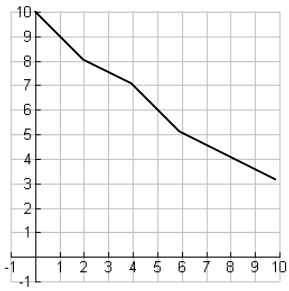
Directions: Sketch a graph to represent how each of the following may look on a graph.

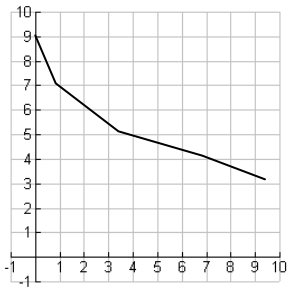
a. time spent on a diet and weight

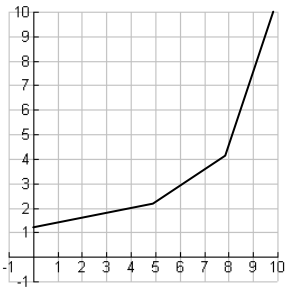
b. age of a car and its value

c. amount of education and salary

Sample graphs include:







Attachment A (continued)
Getting Started Answer Key

The regression equation and coefficient of correlation are given as a guide. Answers for the equation may vary.
Regression Equation: $y = 4.1818x + 5.2$
Coefficient of Correlation: .997692

Sketches for the second part may vary. Parts a and b represent a negative correlation while part c represents a positive correlation (although students may question part c).

Answers for Attachment B:

Page One:

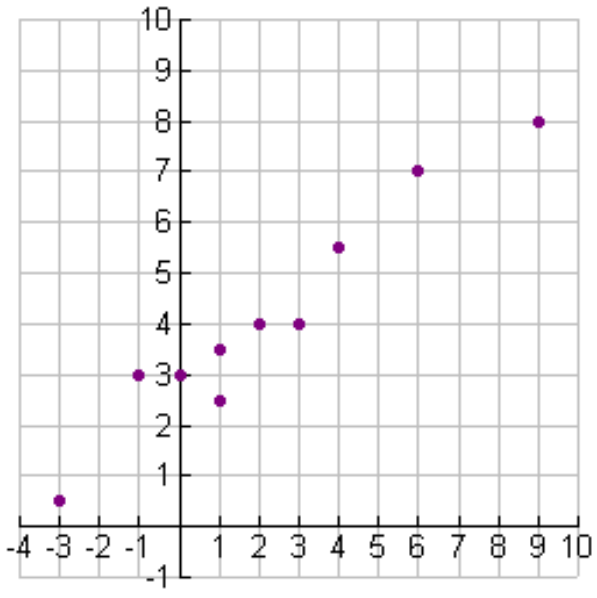
Regression equation: $y = 0.3x + 3.2$
$r = 0.423559$

Page Two:

Regression equation: $y = 0.614051x + 2.74909$
$r = 0.970236$

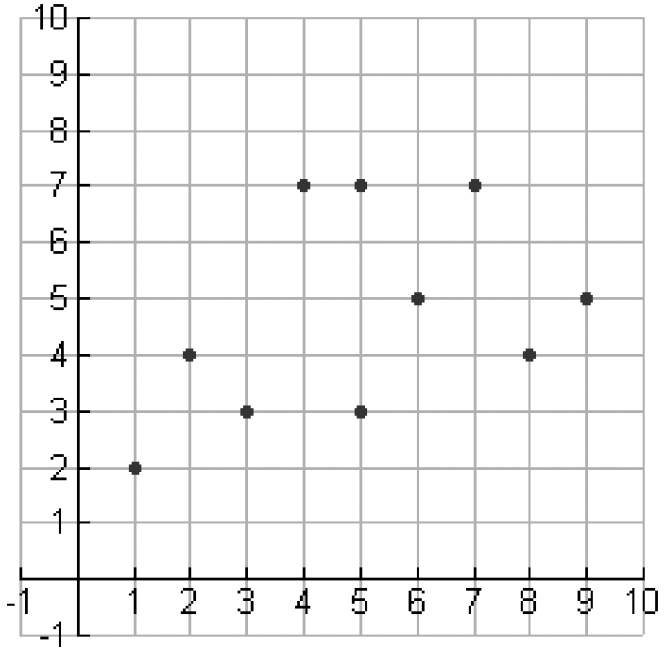
Attachment B
Putting it Together
Page one

x	y
-3	0.5
-1	3
0	3
1	2.5
3	4
4	5.5
6	7
9	8
1	3.5
2	4



Attachment B (continued)
Putting it Together
Page two

x	y
8	4
6	5
9	5
1	2
3	3
5	7
7	7
5	3
2	4
4	7



Attachment C
Group Activity

Name _____ **Date** _____

A

x	y
4	7
3	10
2	5
5	4
1	1
6	12
7	11
0	6

B

x	y
-4	-2
-3	1
-2	1
1	4
0	2
3	6
2	6
-1	3

C

x	y
0	0
1	4
-2	-3
-5	-4
3	6
3	4
5	5
-3	2

D

x	y
0	1
1	-5
-2	1
-5	3
3	-6
3	-3
5	-3
-3	-1

E

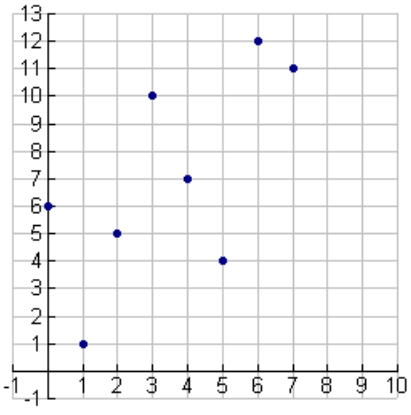
x	y
-6	2
0	5
0	-1
1	-3
-2	-2
-4	1
5	-5
4	-1

F

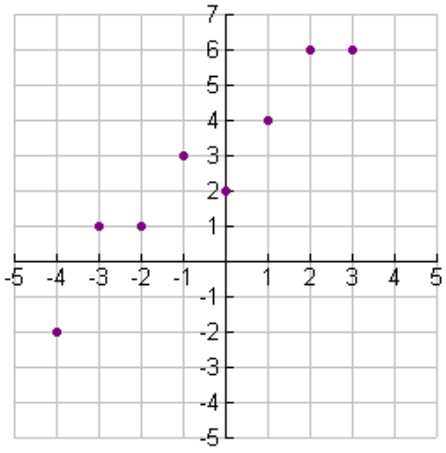
x	y
0	8
-3	10
-1	6
1	6
3	8
-2	8
4	6
6	6

Attachment D
Group Activity Answer Key

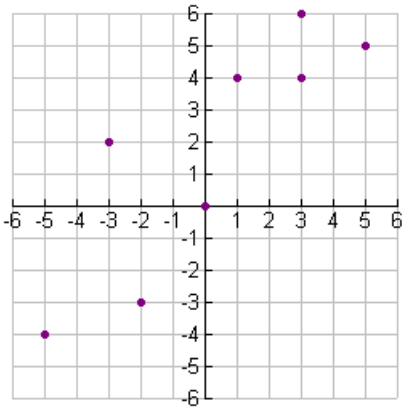
A. $y = 1x + 3.5$
$r = .648$



B. $y = 1.06x + 3.155$
$r = .953$

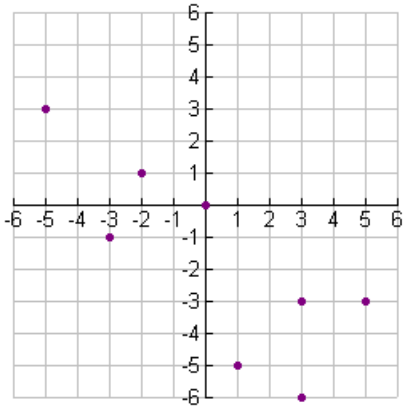


C. $y = .926x + 1.518$
$r = .847$

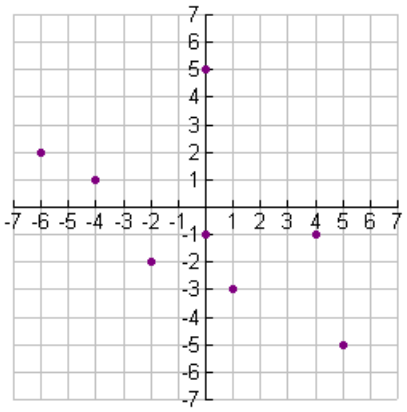


Attachment D (continued)
Group Activity Answer Key

D. $y = -0.706x - 1.574$
$r = -0.787$



E. $y = -0.462x - 0.615$
$r = -0.553$



F. $y = -0.176x + 7.176$
$r = -0.514$